

(19)THE KOREAN INTELLECTUAL PROPERTY OFFICE (KR)  
**(12) Korean Patent Laid-Open Publication (A)**

(11) Korean Patent Laid-Open Publication No.: 1999-0056313  
(43) Korean Patent Laid-Open date: July 15, 1999

---

(21) Application No.: Korean Patent Application No. 10-1997-0076307

(22) Filing Date: December 29, 1997

---

(71) Applicant: Hynix Semiconductor Inc..

(72) Inventor(s): Seong-Hee Park

(74) Attorney(s): Dong-Ho Yoo

---

(54) Title: A SPEECH INTERVAL DETECTION METHOD IN A SPEECH RECOGNITION SYSTEM

---

***Abstract***

The present invention relates to a speech interval detection method in a speech recognition system, for producing an energy of an input signal for each frequency band by using a filter bank, and detecting an accurate speech interval from the input signals by finding correlations of the energies. An RF region of the input signal is emphasized by using a high-pass filter, the input signal with the emphasized RF region is divided into frames with a predetermined size by using the Hamming window, the FFT process is performed for each divided frame, an energy corresponding to each frequency is found by using a triangular filter, correlations of the respective energies are found to produce a speech interval determination index, it is detected that the input signal is a speech signal when the speech interval determination index is greater than a threshold value, and it is detected that the input signal is noise when it is less than the threshold value. Accordingly, the speech interval can be accurately detected from the signal mixed with relatively great noise, and the speech interval of consonants which were difficult to detect because of difficult distinction of the consonants from the noise can also be detected to some degrees.

***REPRESENTATIVE DRAWING***

***Fig. 2***

# (19) 대한민국특허청(KR)

## (12) 공개특허공보(A)

(51) Int. Cl.	(11) 공개번호	특1999-0056313
G10L 5/06	(43) 공개일자	1999년07월15일
(21) 출원번호	10-1997-0076307	
(22) 출원일자	1997년12월29일	
(71) 출원인	현대전자산업 주식회사, 김영환	
	대한민국	
	467701	
	경기도 이천시 부발읍 아미리 산 136-1	
(72) 발명자	박성희	
	대한민국	
	122-010	
	서울특별시 은평구 증산동 2-5	
(74) 대리인	유동호	
(77) 심사청구	없음	
(54) 출원명	음성인식시스템에서의 음성구간 검출방법	

### 요약

본 발명은 음성인식시스템에 있어서, 필터뱅크를 이용하여 입력신호의 각 주파수 대역별 에너지를 산출하고, 이 산출된 에너지들의 상관관계를 파악하여 입력신호로부터 정확한 음성구간을 검출할 수 있도록 한 음성인식시스템에서의 음성구간 검출방법에 것으로, 고역통과필터를 이용하여 입력신호의 고주파 영역을 강조하고, 고주파 영역이 강조된 입력신호를 해밍 윈도우를 사용하여 일정 크기의 프레임으로 세분화하며, 세분화된 프레임 단위로 FFT를 수행한 후 상각형 모양의 필터를 이용하여 각 주파수에 해당되는 에너지를 구하고, 이 구한 각각의 에너지들에 대한 상관관계를 구하여 음성구간 판단지수를 산출한 후, 이 음성구간 판단지수가 임계값 이상인 경우에는 입력되는 신호가 음성신호임을 검출하고, 임계값 이하인 경우에는 잡음신호임을 검출하는 것을 특징으로 하며, 종래의 기술에 비해 상대적으로 큰 잡음이 섞인 신호인 경우에도 음성구간을 정확히 검출할 수 있으며, 잡음과 구별이 어려워 검출이 어려웠던 자음들의 경우에도 어느 정도까지 검출이 가능하게 되는 효과가 있다.

### 대표도

도2

### 명세서

#### 도면의 간단한 설명

도 1은 일반적인 음성인식시스템의 블록 구성도,

도 2는 본 발명에 의한 음성인식시스템에서의 음성구간 검출 흐름도,

도 3은 본 발명에 의한 필터뱅크를 나타낸 도면,

#### <도면의 주요부분에 대한 부호의 설명>

1 : 음성구간 검출부                      2 : 특징계수 추출부

3 : 음성 인식부                          4 : 데이터 베이스

5 : 후처리부

#### 발명의 상세한 설명

##### 발명의 목적

##### 발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 인간의 음성을 인식하기 위한 음성인식시스템에 있어서, 필터뱅크(Filter Bank)를 이용하여 입력신호의 각 주파수 대역별 에너지를 산출하고, 이 산출된 에너지들의 상관관계를 파악하여 입력신호로부터 정확한 음성구간을 검출할 수 있도록 한 음성인식시스템에서의 음성구간 검출방법에 관한 것이다.

일반적으로, 인간의 음성과 같은 자연음을 인식하고자 하는 음성인식시스템은 도 1에 도시된 바와 같이 입력되는 신호로부터 음성구간을 검출하는 음성구간 검출부(1)와, MFCC 계수를 사용하여 상기 음성구간 검출부(1)에서 검출된 음성구간으로부터 그 특징을 추출하는 특징계수 추출부(2)와, HMM(Hidden Markov Model)과 VMS VQ(Variable Multi-Section Vector Quantization) 알고리즘을 이용하여 음성신호를 인식하는 음성 인식부(3)와, 음성신호에 의해 학습된 단어모델 파라미터가 저장되어 있는 데이터 베이스(4)와, 상기 음성 인식부(3)에서 인식된 음성신호에 대해 실효성을 판단하여 인식되는 단어를 출력하는 후처리부(5)로 구성된다.

상기와 같이 구성된 음성인식시스템에 있어서, 입력되는 신호로부터 정확한 음성구간을 검출하는 것은 음성인식시스템의 전처리 부분으로, 시스템의 성능을 좌우하는 전제조건으로서 매우 중요한 작업이다.

그러나, 잡음이 많은 외부 환경에서는 잡음성분으로 인하여 정확한 음성구간을 검출하기가 쉽지 않다.

이에 따라, 종래에는 입력신호의 각 프레임(Frame)의 최대값 에너지 또는 평균 에너지나 영교차율(Zero Crossing Rate) 등의 정보를 이용하여 배경잡음으로부터 산출한 임계값과 비교함으로써 음성구간을 검출하였다.

그러나, 이러한 검출방법은 배경잡음이 작을 경우에는 비교적 음성구간을 잘 검출하지만, 배경잡음이 상대적으로 클 경우에는 음성구간 검출 성능이 현저하게 떨어지게 되는 문제점이 있었다.

또한, 잡음성분과 쉽게 구별되지 않는 시뮬,지뮬,티뮬,키뮬 등의 자음에 대해서는 아예 음성구간에 포함되어 있지 않게 되어 검출이 불가능한 문제가 있었다.

#### 발명이 이루고자 하는 기술적 과제

본 발명은 상기와 같은 문제점을 해결하기 위해 안출한 것으로서, 그 목적은 필터뱅크를 이용하여 입력신호의 각 주파수 대역별 에너지를 산출하고, 이 산출된 에너지의 상관관계를 통해 음성신호와 잡음신호를 구별하여 입력신호로부터 음성구간만을 효과적으로 검출할 수 있도록 한 음성인식시스템에서의 음성구간 검출방법을 제공하는 데에 있다.

이러한 목적을 달성하기 위한 본 발명의 음성인식시스템에서의 음성구간 검출방법은, 고역통과필터를 이용하여 입력신호의 고주파 영역을 강조하고, 고주파 영역이 강조된 입력신호를 해밍 윈도우를 사용하여 일정 크기의 프레임으로 세분화하며, 세분화된 프레임 단위로 FFT를 수행한 후 삼각형 모양의 필터를 이용하여 각 주파수에 해당되는 에너지를 구하고, 이 구한 각각의 에너지들에 대한 상관관계를 구하여 음성구간 판단지수를 산출한 후, 이 음성구간 판단지수가 임계값 이상인 경우에는 입력되는 신호가 음성신호임을 검출하고, 임계값 이하인 경우에는 잡음신호임을 검출하는 것을 특징으로 한다.

#### 발명의 구성 및 작용

이하, 도 2를 참고하여 본 발명에 의한 음성인식시스템에서의 음성구간 검출방법을 상세히 설명한다.

먼저, 음성인식시스템으로 입력되는 입력신호를 아래 수학적 식 1과 같은 고역통과필터를 이용하여 고주파 영역을 강조한다(S10).

수학적 식 1

$$H(Z)=1-aZ^{-1}, \quad (0.9 \leq a \leq 1.0)$$

이어, 상기 단계(S10)에서 고주파 영역이 강조된 입력신호를 아래 수학적 식 2와 같은 해밍(Hamming) 윈도우(Window)를 사용하여 10ms씩 중첩하여 20ms 크기의 프레임으로 세분화(Blocking)한다(S11).

수학적 식 2

$$W(n)=0.54-0.46\cos\left(\frac{2\pi n}{N-1}\right)$$

여기서,  $0 \leq n \leq N-1$ ,  $N$ : 프레임당 샘플수

그리고 나서, 상기 단계(S11)에서 세분화된 각 프레임 단위로 256-포인트(point) FFT(Fast Fourier Transform)를 수행하면(S12), 이 FFT 결과를 도 3에 도시된 바와 같은 삼각형 모양의 필터, 즉 필터뱅크를 이용하여 각 주파수에 해당되는 에너지를 구한다(S13).

이어서, 수학적 식 3을 이용하여 상기 단계(S13)에서 필터뱅크를 거친 신호를 미리 준비된 배경잡음신호에 대하여 상관관계 계수를 구한다(S14).

수학적 식 3

$$\Phi_n(i,k)=\sum_{m=0}^{N-1+P} x_n(m-i)S_n(m-k)$$

여기서,  $1 \leq i \leq P$ ,  $0 \leq k \leq P$ ,  $P$ : 필터뱅크의 계수,  $n$ :  $n$  번째 프레임

이후, 상기 단계(S14)에서 구한 상관관계 계수로부터 수학적 식 4와 같이 음성구간 판단지수를 산출하고(S15), 이 음성구간 판단지수를 이용하여 음성신호 여부를 판단한다(S16).

수학적 식 4

$$\alpha_n = \sum_{i=1}^P \sum_{k=0}^P \Phi_n(i,k)$$

즉, 상기 음성구간 판단지수가 임계값 이상인 경우에는 입력되는 신호가 음성신호임을 검출하고(S17), 임계값 이하인 경우에는 잡음신호임을 각각 검출하게 된다(S18).

#### 발명의 효과

이상, 상기 설명에서와 같이 본 발명은 주파수 영역에서의 각 주파수 대역의 에너지 값들의 상관관계를 산출하여 이를 근거로 하여 입력신호로부터 음성신호를 분리해냄으로써, 종래의 기술에 비해 상대적으로 큰 잡음이 섞인 신호인 경우에도 음성구간을 정확히 검출할 수 있으며, 잡음과 구별이 어려워 검출이 어려웠던 자음들의 경우에도 어느 정도까지 검출이 가능하게 되는 효과가 있다.

#### (57) 청구의 범위

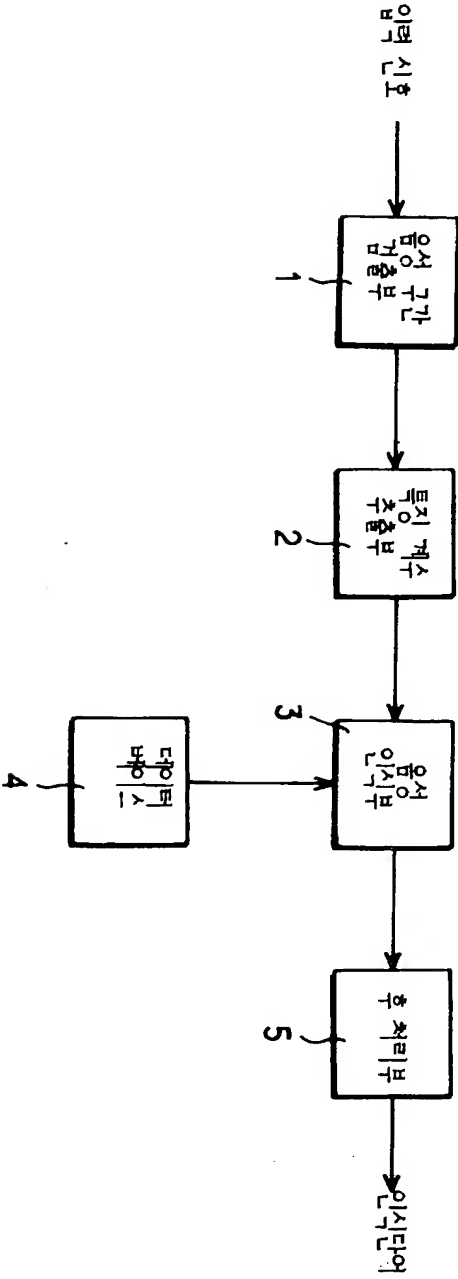
##### 청구항 1.

음성인식 시스템에서의 음성구간을 검출하는 방법에 있어서,

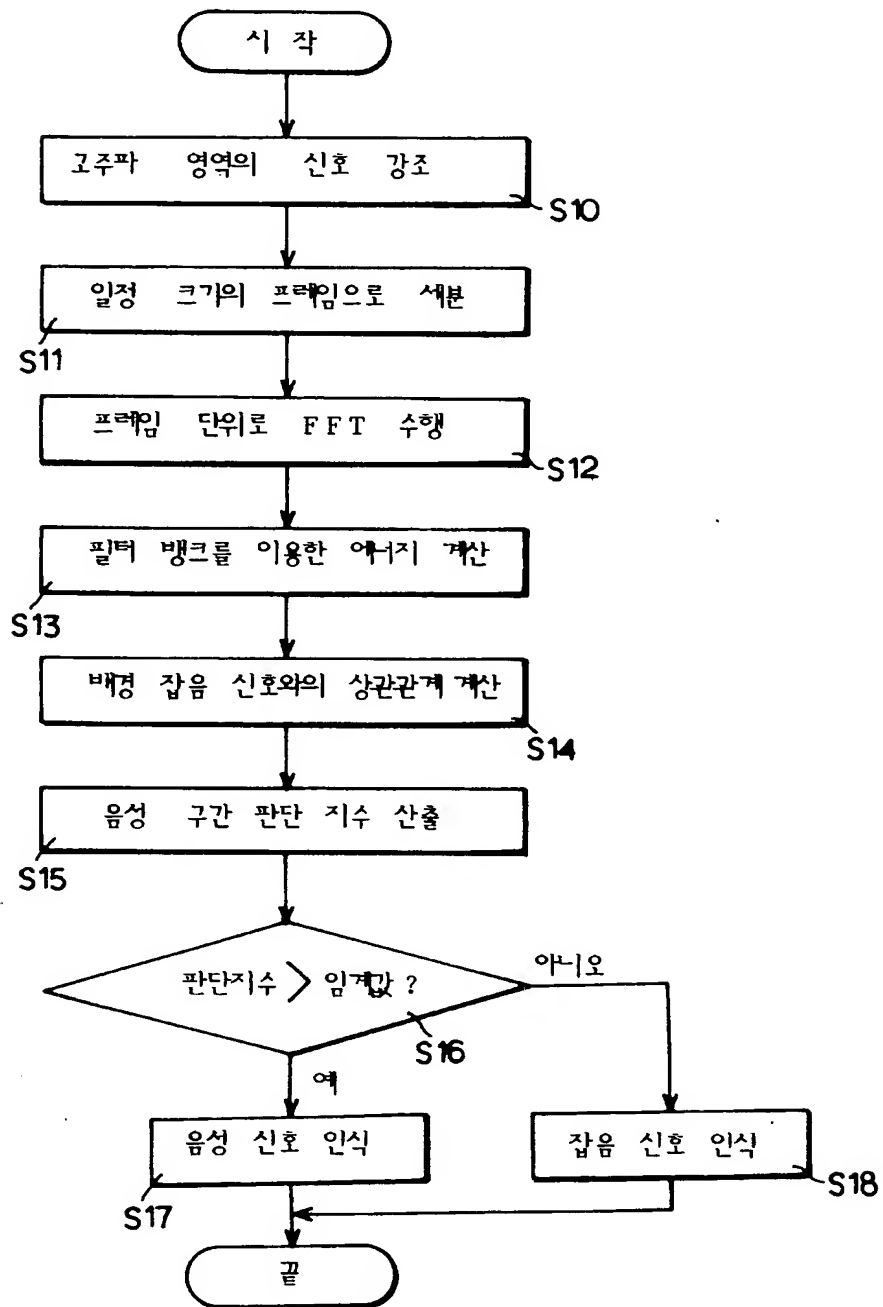
고역통과필터를 이용하여 입력신호의 고주파 영역을 강조하는 제1단계와, 상기 제1단계에서 고주파 영역이 강조된 입력신호를 해밍 윈도우를 사용하여 일정 크기의 프레임으로 세분화하는 제2단계와, 상기 제2단계에서 세분화된 입력신호를 각 프레임 단위로 FFT를 수행하는 제3단계와, 상기 제3단계에서 수행된 FFT 결과를 필터뱅크를 이용하여 각 주파수에 해당되는 에너지를 구하는 제4단계와, 상기 제4단계에서 필터뱅크를 거친 입력신호를 미리 준비된 배경잡음신호에 대해 상관관계 계수를 구하는 제5단계와, 상기 제5단계에서 구한 상관관계 계수를 이용하여 음성구간 판단지수를 산출하는 제6단계와, 상기 제6단계에서 산출된 음성구간 판단지수가 임계값 이상인 경우에는 입력되는 신호가 음성신호임을 검출하고, 임계값 이하인 경우에는 잡음신호임을 검출하는 제7단계로 이루어지는 것을 특징으로 하는 음성인식시스템에서의 음성구간 검출방법

도면

도면 1



도면 2



도면 3

